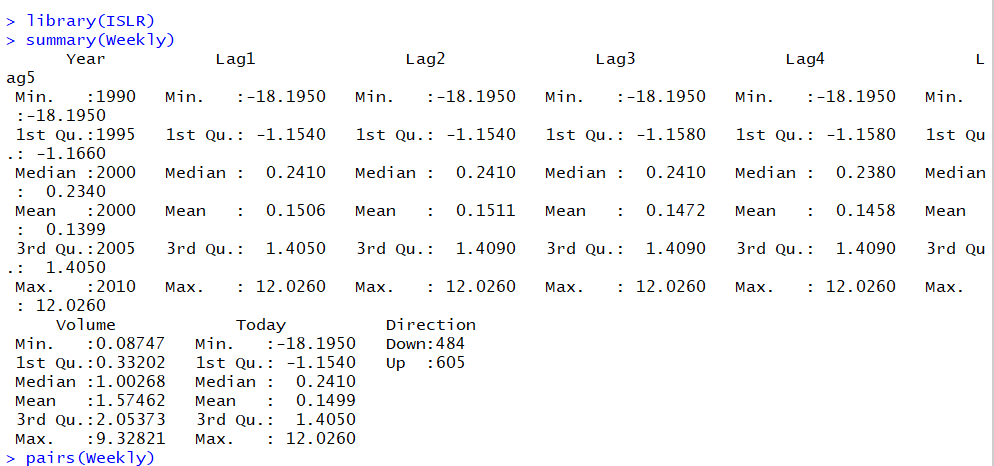
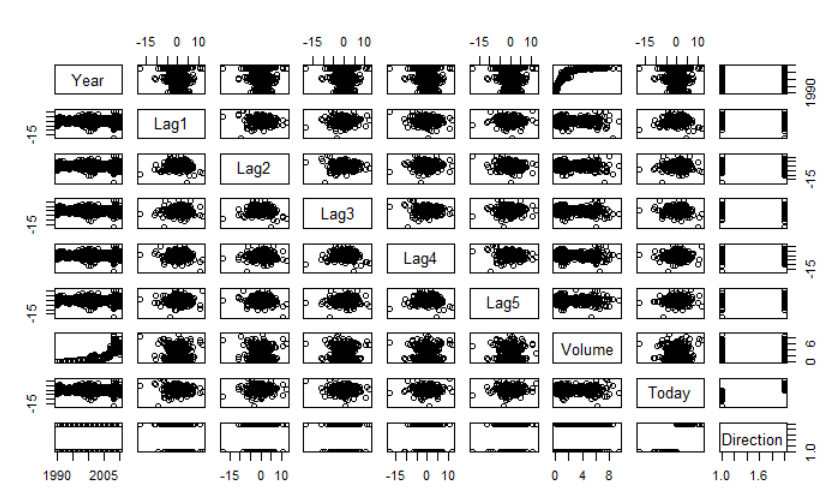
**ISL LAB ASSIGNENT-3**

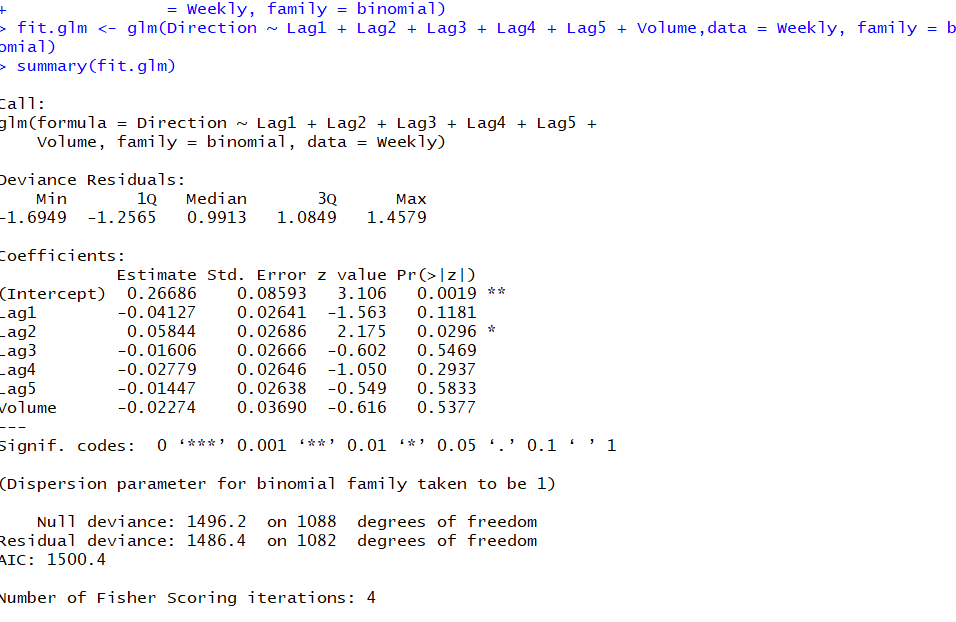
**Tejaswi Ayyadapu-16278799**



2(a)



2(b)



Lag2 has least p value so we can say that lag2 is statistically significant.

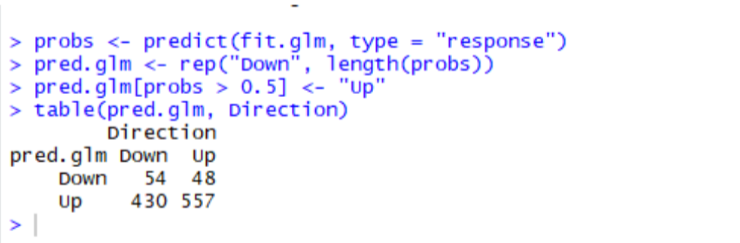
2(c)

probs <- predict(fit.glm, type = "response")

pred.glm <- rep("Down", length(probs))

pred.glm[probs > 0.5] <- "Up"

table(pred.glm, Direction)



We may conclude that the percentage of correct predictions on the training data

is (54+557)/1089 which is equal to 56.1065197%. In other words 43.8934803% is the

training error rate, which is often overly optimistic. We could also say that for weeks when

the market goes up, the model is right 92.0661157% of the time (557/(48+557)). For weeks

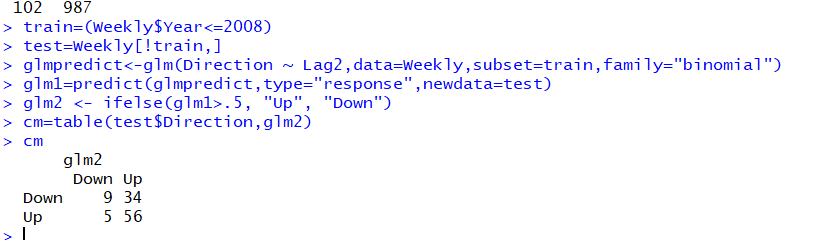
when the market goes down, the model is right only 11.1570248% of the time

(54/(54+430)).

There are a predominance of Up prediction. The model predicts well the Up direction, but it

predict poorly the Down direction.

2(d)

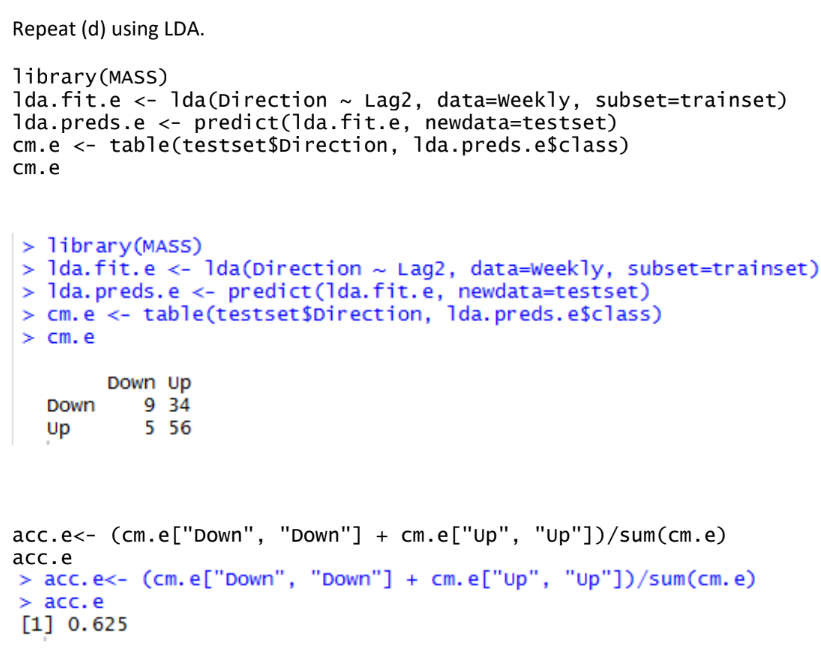


> acc.d <- (cm["Down", "Down"] + cm["Up", "Up"])/sum(cm)

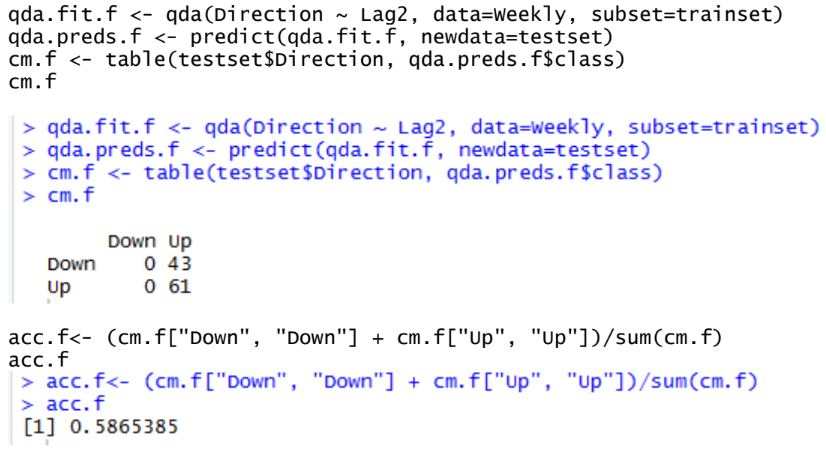
> acc.d

[1] 0.625

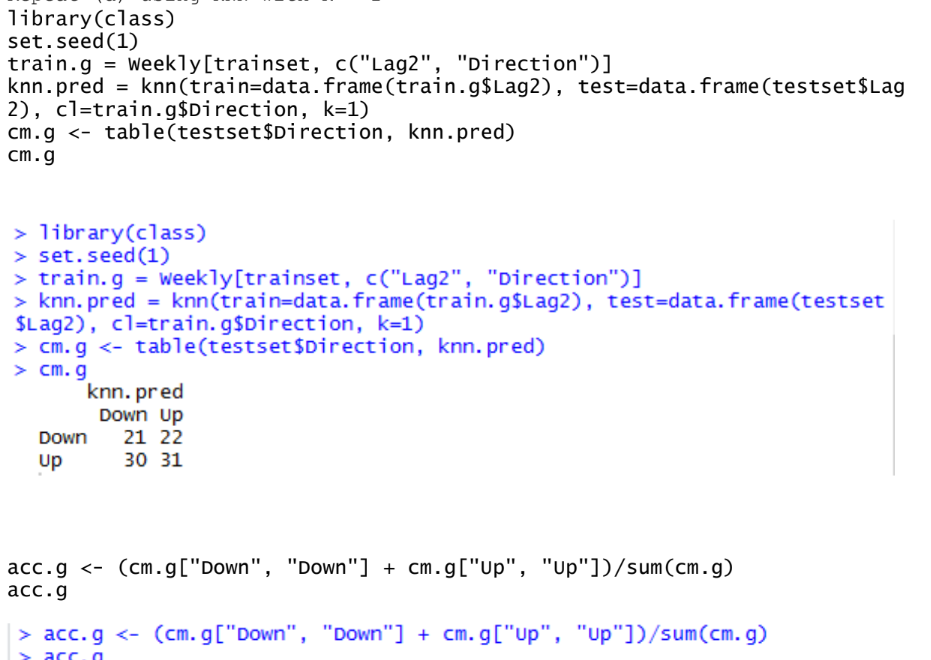
2(e)



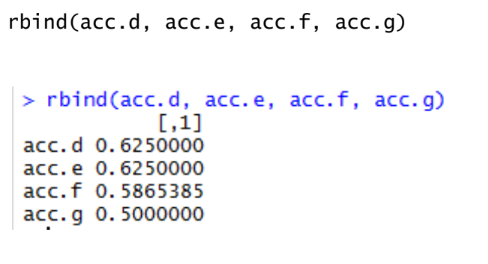
2(f)



2(g)



2(h)



best results on this data are logistic regression and LDA

2(i)

> set.seed(1)

>

> results <- data.frame(k=1:50, acc=NA)

> for(i in 1:50){

+ knn.pred = knn(train=data.frame(train.g$Lag2), test=data.frame(testset$Lag2), cl=train.

g$Direction, k=i)

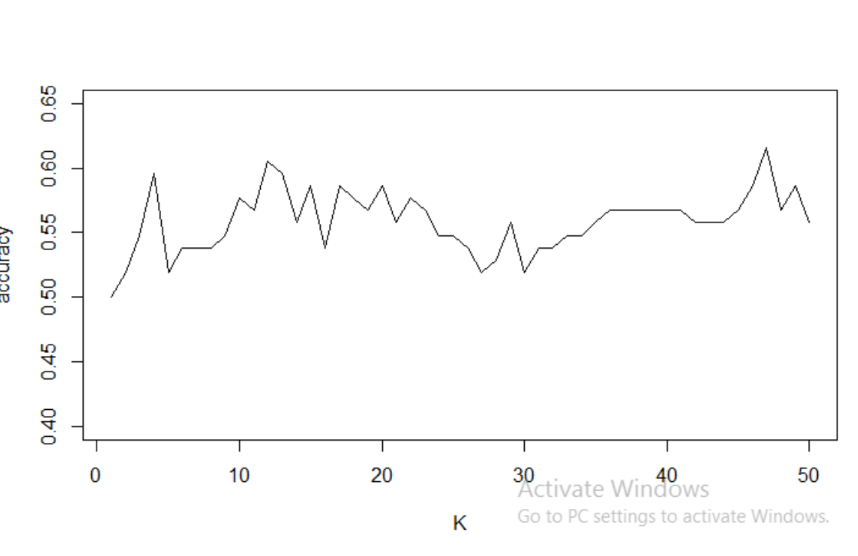
+ cm <- table(testset$Direction, knn.pred)

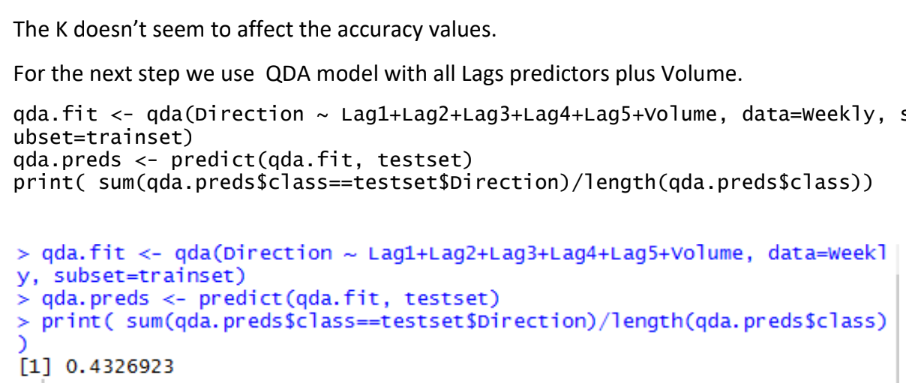
+ acc <- (cm["Down", "Down"] + cm["Up", "Up"])/sum(cm)

+ results$acc[i] <- acc

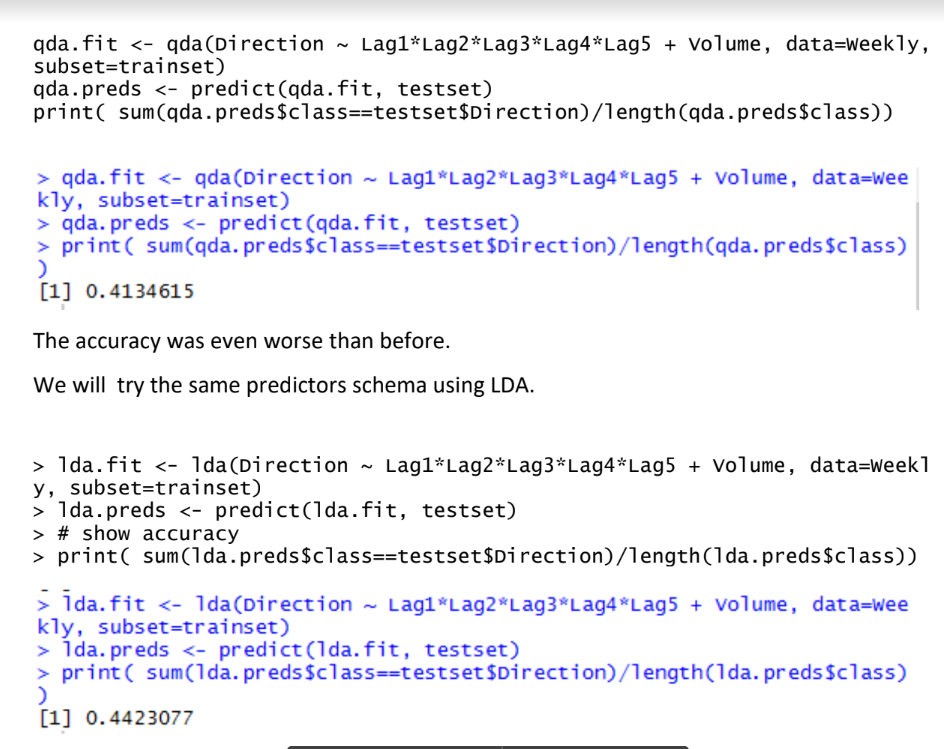
+ }

> plot(x=results$k, y=results$acc, type="l", xlab="K", ylab="accuracy", ylim=c(.4,.65))



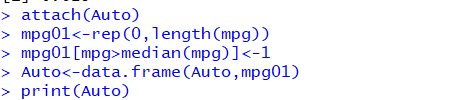


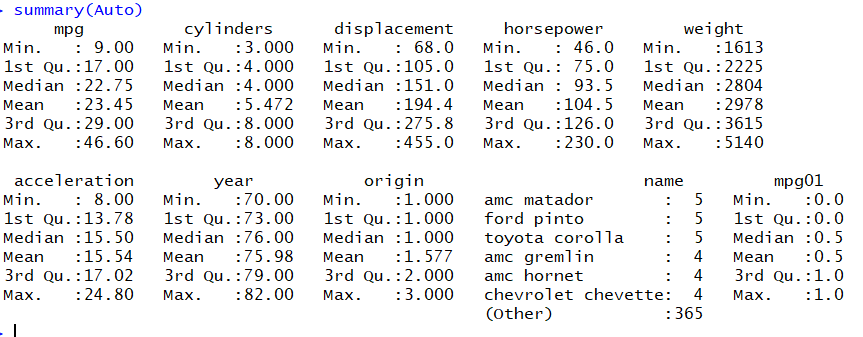
We use again the QDA model with interactive variables between all Lag predictors.



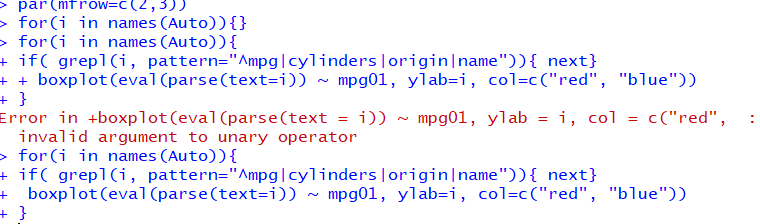
The LDA performance is similar with QDA

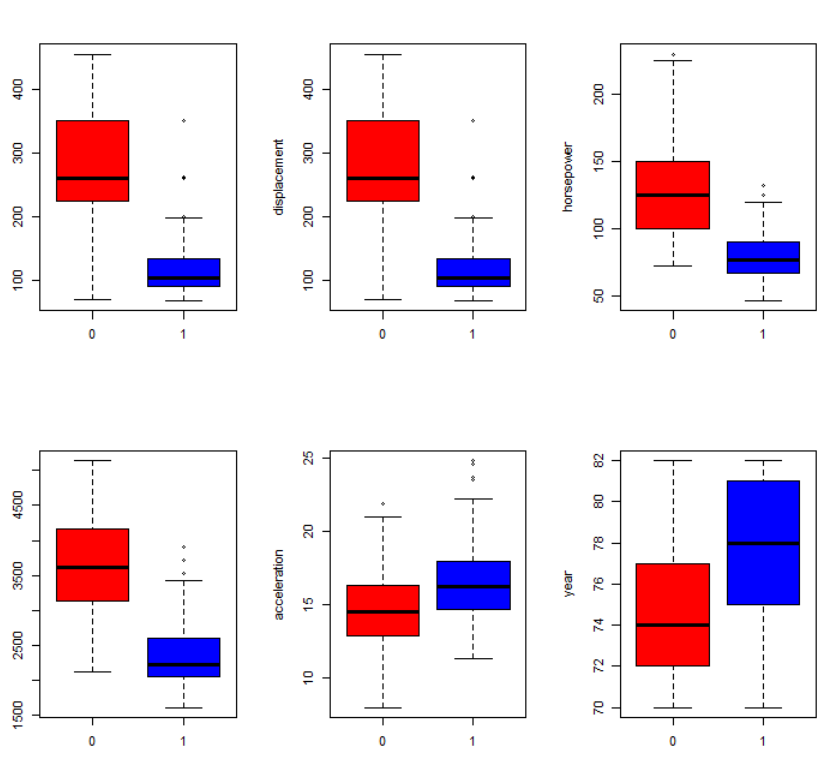
3.a.



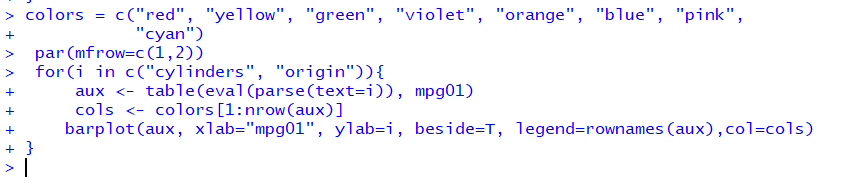


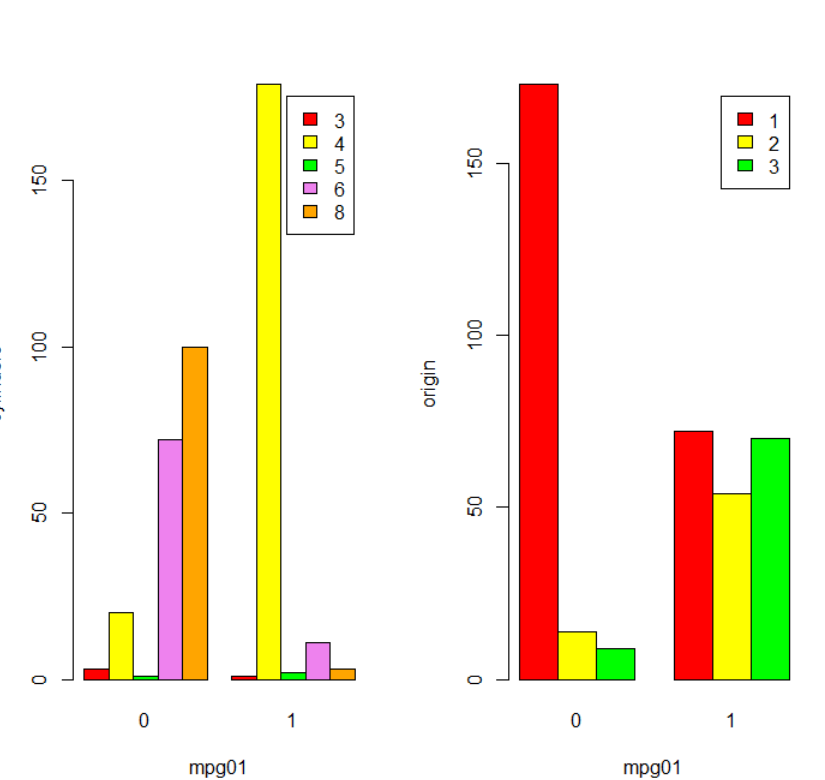
3(b)





3(b)

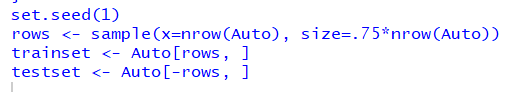




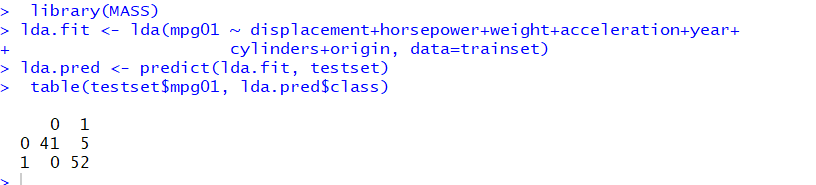
At the bar plots, cylinders and origin also show relation with mpg01. For instance, on

dataset cars of lower mpg are majority from origin 1, which is American.

3(c)



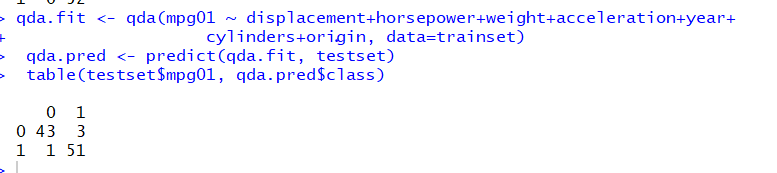
3(d)



round(sum(lda.pred$class!=testset$mpg01)/nrow(testset)\*100,2)

5.1

3(e)



round(sum(qda.pred$class!=testset$mpg01)/nrow(testset)\*100,2)

4.08

3(f)

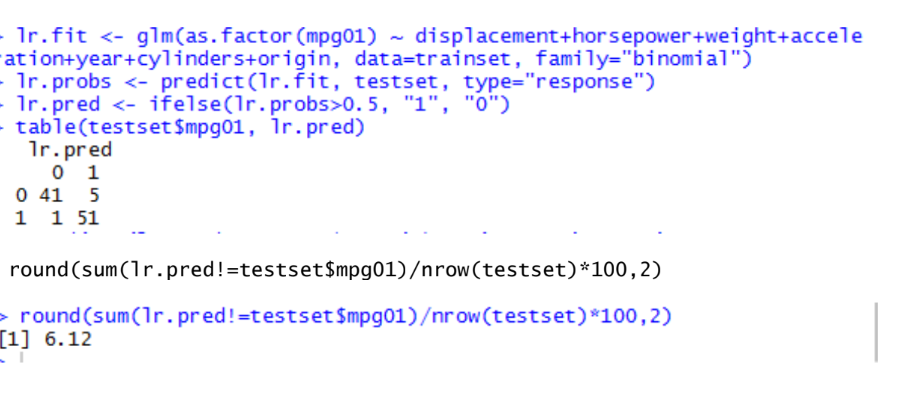
lr.fit <- glm(as.factor(mpg01) ~ displacement+horsepower+weight+accelera

tion+year+cylinders+origin, data=trainset, family="binomial")

> lr.probs <- predict(lr.fit, testset, type="response")

> lr.pred <- ifelse(lr.probs>0.5, "1", "0")

> table(testset$mpg01, lr.pred)



3(g)

